DOI:
10.1111/bju.13710

Document Version
Peer reviewed version

Received Date : 01-Aug-2016
Accepted Date : 04-Nov-2016
Article type : Original Article
Article category: Robotics and Laparoscopy

Competency Based Training in Robotic Surgery: Benchmark Scores for Virtual Reality Robotic Simulation

N Raison MBBS[1,2], K Ahmed PhD[1], N Fossati MD[3], N Buffi MD[3], A Mottrie PhD[5], P Dasgupta MD[1], H van der Poel PhD[6]

[1]MRC Centre for Transplantation, Faculty of Life Sciences & Medicine, King's College London, UK
[2]The London Clinic, London, UK
[3]IRCCS Ospedale San Raffaele, Milan, Italy
[4]Humanitas Research Hospital, Milan Italy
[5]OLV Hospital, Aalst, Belgium
[6]Netherlands Cancer Institute, Amsterdam, The Netherlands

Corresponding Author
Dr Nicholas Raison
Address: Division of Transplantation Immunology & Mucosal Biology, King's College London, 5th Floor Tower Wing, Guy's Hospital, London SE1 9RT
Tel: 0207 188 5906
Fax: 0207 188 0944
Email: Nicholas.raison@kcl.ac.uk

Keywords: curriculum, hands-on-training, robotic surgery, virtual reality simulation training

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bju.13710

Abstract

Objectives

To develop benchmark scores of competency for use within a competency-based virtual reality (VR) robotic training curriculum.

Subjects and Methods

This longitudinal, observational study analysed results from 9 EAU hands-on-training courses in VR simulation. 223 participants ranging from novice to expert robotic surgeons completed 1565 exercises. Competency was set at 75% of the mean expert score. Benchmark scores for all general performances metrics generated by the simulator were calculated. Assessment exercises were selected by expert consensus and through learning curve analysis. Three basic skill and two advanced skill exercises were identified.

Results

Benchmark scores based on expert performance offered viable targets for novice and intermediate trainees in robotic surgery. Novice participants met the competency standards for most basic skill exercises however advanced exercises were significantly more challenging. Intermediate participants performed better across the seven metrics but still fell short of the benchmark standard in the more difficult exercises.

Conclusion

Benchmark scores derived from expert performances offer relevant and challenging scores for trainees to achieve during VR simulation training. Objective feedback allows both participants and trainers to monitor educational progress and ensures that training remains effective. Furthermore, the well-defined goals set through benchmarking offer clear targets for trainees and enable training to move to a more efficient competency based curriculum.

Introduction

Simulation training has undergone a significant period of expansion in recent years. It is increasingly recognised that simulation training is integral to the surgical curriculum[1]. Virtual reality (VR) simulators offer a realistic imitation of robotic surgery allowing training outside the highly stressful and expensive operating room environment [2,3]. Not only is this safer for patients but it allows training to be driven by learning needs rather than dictated by caseload and patient availability[4].

VR simulators have been extensively validated, demonstrating their potential usefulness for training within structured curricula [5,6]. Yet while assessment and feedback are both key components of effective learning, they remain poorly integrated into simulation training programmes if at all [7].

The validated EAU Hands-On-Training (HOT) courses deliver VR training to familiarise participants with the basic and advanced skills required for robotic surgery[8]. Participants receive practical tuition from expert surgeons but no formal assessment of a participants' progression during the course is made.

Aims

To develop objective benchmark scores of competency for use during VR robotic surgical training.

Methods

HOT courses are held regularly during EAU conferences and symposia. Each course, which last for 90min, provides VR simulation training using the Mimic dV Trainer (Mimic Technologies, Inc, Seattle, WA) mentored by robotic expert surgeons. As introductory training, the course aims to teach basic robotic skills. Course participation is open to all conference delegates regardless of experience or training. A prospective, observational study was conducted recruiting candidates from nine robotic VR simulator HOT courses between March 2013 and September 2015.

Demographic details and surgical experience were collected using a pre-course questionnaire distributed to all participants. Participants' simulation exercise results were retrieved electronically from the simulators and compiled on a database.

Across all nine courses, 16 different exercises were completed. For each exercise the simulator's inbuilt algorithms calculated a variety of performance metrics. These can be divided into general performance scores, applicable to all exercises, (Time to Complete, Economy of Motion, Master Working Space, Instruments Out of View, Excessive Force, Instrument Collision) or task specific scores (Blood Loss, Broken Vessels, Misapplied Energy Time, Missed Target, Dropped instruments). An overall summary score is generated based on the individual's performance across all metrics. Analysis was limited to general performance metrics and overall score so that all benchmark scores were applicable to all exercises.

To comprehensively assess a participant's performance, a range of test exercises were selected encompassing basic and advanced skills. Suitable exercises were selected by expert agreement based on task complexity, skill focus area of each exercise and analysis of the previous HOT course participant results. For valid analysis, only exercises performed at least 80 times were included. Basic tasks were required to assess performance in generic, fundamental robotic skills. The majority of participants were expected to be able to complete these exercises competently with relatively homogenous exercise scores. In contrast, advanced skill exercises required more complex, task specific, robotic skills. The greater challenge posed by these exercises was expected to result in greater variability and overall lower scores.

Benchmark scores were set using a criterion-referenced method based on expert scores. Given that the HOT course aims to provide initial training in basic robot skills, benchmark scores were required to identify participants that had achieved a minimum level of competency rather than proficiency. Use of mean expert scores to gauge proficiency has been previously demonstrated[9]. Potential benchmark standards for competency were set at 60%, 75% and 90% of the mean expert score. These were modelled against participant outcome data to identify an appropriate standard (Figure 1). A competency standard of 75% of mean expert score was found to set a suitable standard, based on the performance of novice (no robotic surgical experience), intermediate (1-74 robotic procedures performed) and expert participants (>75 robotic procedures performed). Retrospective analysis of the time to complete metric showed appropriate competency rates in comparison to 90% and 60% standards (Supplementary Table S1).

Expert robotic surgeons were defined as having performed over 75 robotic cases independently. Determining clinical proficiency is difficult given learning curves remain both procedure and surgeon specific. This level of experience was selected of the basis of the available literature on robotic surgical learning curves[10]. Data from both expert performances during HOT courses and from the worldwide Mimic score database was combined to produce an overall expert score for each exercise. Experts from the Mimic database were recruited from 6 institutions in the US, France and Sweden.

To test the suitability of the benchmark scores, comparisons were made between participants with differing levels of robotic surgical experience. Participants were divided into novice (no robotic surgical experience) and intermediate groups (1-74 robotic procedures performed). Their respective performances were compared against the benchmark criteria and expert performance scores.

Statistical Analysis

Initial data analysis demonstrated non-Gaussian distribution of scores with uniformly positively skewed data. To generate a normally distributed data set, logarithmic transformation was performed[11]. All further analysis was performed on this log-normal data set. For metric scores in each case the geometric mean was calculated.

All calculations were performed using SPSS version 22.0 (Armonk, NY, IBM Corp). Graphs were created using Prism version 6.0 (GraphPad Software, Inc. La Jolla, California, USA).

Results

Demographics

223 participants completed 1565 exercises during the nine HOT courses. Demographic details and surgical experience are shown in Table 1. The HOT course cohort was composed of resident and attending urologists from 21 countries. Overall robotic experience was low. Residents had a mean ± standard deviation experience of assisting in 26.8 ± 62.0 cases and performing 1.4 ± 3.8 cases whilst attending urologists had experience of 19.3 ± 44.5 cases and 7.3 ± 33.1 cases respectively. Robotic simulation experience was equally low with 50.1% having no prior simulation experience. Previous analysis has shown that the main factors influencing the overall score of participants were age and prior robotic experience[12]. Each participant completed a mean of 7 exercises during the HOT course.

Identification of Assessment Tasks

Nine simulator exercises were performed >80 times (Pick & Place, Camera Targeting 1, Energy Dissection 1, Suture Sponge, Peg Board 1, Camera Targeting 2, Peg Board 2, Ring Walk 2). From these, five were selected as suitable assessment exercises; three basic level exercises and two advanced level exercises. The three basic exercises (Pick and Place, Camera Targeting 1, Peg Board 1) tested fundamental robot skills including endowrist manipulation, clutching, 3-D vision and camera control. As predicted results of participant performances showed relatively homogenous results with a major proportion of participants achieving high overall scores. Yet interestingly all three exercises had a pronounced dichotomy in scores, with a clear division between trainees (Supplementary Figure S2). Closer analysis revealed that while most participants started with these basic skill tasks, there was little repetition even following poor performance and hence no scope for development.

Suture Sponge and Thread the Rings 1 were selected as advanced assessment tasks. Both exercises assessed the more complex skill of suturing requiring needle driving in addition to competent execution of basic robotic surgical skills. Performance analysis showed far greater variability in scores and overall lower scores as expected with the more complicated tasks.

Development of Benchmark Criteria

The benchmark score for the minimum necessary standard to be achieved was set as 75% of the geometric mean expert performance. Scores were calculated for all general performance metrics (Time to Complete, Economy of Motion, Master Workspace, Instruments Out of View, Instrument Collisions, Excessive Force and Overall Score) for each of the five exercises. Mean participant scores were compared to the benchmark and expert scores in each case (Figure 1). For basic tasks (Pick & Place, Camera Targeting 1, Peg Board 1), participant mean scores closely correlated with the benchmarks. Scores for the two advanced tasks (Thread the Rings 1, Suture Sponge) were more disparate. The key metrics for basic skills tasks were Time to Complete, Economy of Motion and Instruments Out of View. In contrast Excessive Force and Instruments Out of View were only effective measures of competency for advanced tasks. Overall score, as a cumulative score of the other performance metrics, provided a gross summary however its usefulness in assessing competence was limited especially in basic exercises where most participants met the competency standard (Figure 1c). Mean participant score met or exceeded the benchmark in all but one exercise. Master Working Space (Figure 1e) was found not to be a suitable performance indicator with all participants exceeding the benchmark criteria.

Comparative Benchmark Assessment of Novice and Intermediate Participants

Benchmark scores offer appropriate targets for both novice and intermediate participants (Figure 2 & 3). The majority of novice participants demonstrated competency in the basic tasks however fewer were able to meet the benchmark score for the advanced skills assessment tasks. As expected intermediate participants performed better in the majority of tasks across the six metrics but still fell short of the benchmark scores in the more difficult exercises.

Time to Complete (Figure 2a)

The increase in the complexity of the exercises is clearly reflected in the Time to Complete metric with a progressive rise in the benchmark standard. Intermediate participants demonstrated competency in 3/5 exercises but a proportionally greater skills gap is seen with novice candidates.

Economy of Motion (Figure 2b)

Economy of motion exhibits a similar progression in difficulty across the five exercises. Benchmark scores become increasingly challenging for novice candidates unlike intermediate candidates who remain close to the benchmark standard and even surpass it in Suture Sponge.

Excessive Force (Figure 2c)

Excessive Force offers a greater challenge for inexperienced participants. Experts are able to maintain low scores across both basic and advanced exercises. In contrast, scores for both novices and intermediates deviate markedly with Thread the Rings 1 and Suture Sponge. Uniformly low scores in the basic skill tasks limits its application in these exercises.

Instruments Out of View (Figure 3a)

Whilst this metric does not adhere directly to the previously seen pattern of rising task complexity most notable is the poor participant scores during Camera Targeting 1. Again intermediate and novice participants perform significantly worse during the two more complex tasks.

Instrument Collisions (Figure 3b)

Like Excessive Force, basic task scores were uniformly low preventing effective differentiation of participants. With the increasing difficulty of the tasks, the participants' scores rise exponentially compared to expert scores.

Overall Score (Figure 3c)

As a weighted composite score, overall scores would be expected to be equal across all tasks, demonstrated by relatively uniform expert scores. All participants achieved competency in the basic skill exercises although there was a greater variation between intermediate and novice scores. In contrast novice participants failed to meet the benchmark for either of the advance exercises and intermediates only just met the standard for Suture Sponge.

Discussion

The application of simulation-based training in surgery using artificial but realistic learning environments continues to grow. Extensive research has been undertaken in confirming the educational potential of surgical simulations. Particularly within robotic surgery, VR simulators have been comprehensively validated[13]. In contrast assessment and standard setting for simulation-based training has been largely disregarded.

Historically surgical education has been based on repeated practice, with learning both contingent on and judged by case experience. Whilst concepts such as minimum case numbers remain ubiquitous, the value of competency-based training through simulation is increasingly being recognised. Yet potential benefits of simulation training remain dependent on the trainee and their ability to learn. Factors such as cognitive ability, motivation, perceived utility of training and self-confidence, can account for a significant proportion of the variation seen in training outcomes[14]. Hence 'real-world' results of simulation training will not necessarily match those from a highly focussed trial setting. Objective assessment is needed to confirm to both trainers and trainees that the educational objectives of a training programme have been met. We have demonstrated that benchmark scores, based on expert performance, set relevant targets for participants irrespective of their experience during HOT courses. Competency standards for basic tasks (Pick & Place, Camera Targeting 1, Peg Board 1) were achievable by the majority of participants. In contrast, novice mean performance scores fell below the benchmark standard for all advanced tasks. Similarly, the intermediate candidates also failed to reach the standard for the majority of exercises.

VR simulators provide a wide range of data on an individual's performance. Yet without defined benchmarks, results remain abstract and unrelated to clinical performance. Determining competence is key to effective training but establishing credible, appropriate cut off scores remains challenging [15]. Competency assessment demands an evidence based approach against impartial standards. Lack of external standards excludes the use of norm-referenced benchmarks and the traditional use of expert opinions to set standards introduces the potential for subjectivity and bias[16].

Criterion-referenced benchmark standards based on expert scores not only provide considerable face validity but offer an objective and clinically relevant marker for assessment. Criterion-based assessment has been used in the past albeit infrequently with the mean expert score predominantly used as the standard[17,18] with the average expert benchmark representing proficiency[18] or "optimal performance"[17]. In contrast, HOT courses aim to provide basic training so that participants gain competency in basic robotic surgical skills rather than reaching proficiency. For this reason, the benchmark criterion was set as 75% of the mean expert score.

Metric based benchmarks also offer the benefits for trainees. Specific goals help motivate participants and the immediate feedback will highlight skill domains that require improvement, aiding reflection and deliberate practice[19].

Specifying benchmark scores for all generic performance metrics across different robotic skills exercises permits stepwise training. Division of training into sequential tasks of increasing difficulty mirrors the process of motor skill acquisition [20]. Trainees initially gain familiarity with robotic controls and basic skills, such as clutch control, camera control and endowrist manipulation. Subsequently, trainees apply these skills to more advanced tasks such as knot tying or suturing. This involves both refining their basic skills alongside learning such advanced techniques. Using benchmark criteria to govern progression ensures that course participants achieve competency in basic skills before progressing to more complex tasks. In contrast, unstructured training with progression regardless of scores risks poorer training outcomes. This may explain the poor progression of previous participants during HOT courses (Supplementary Figure S2).

On the basis of this study, the authors propose a modular training programme for VR simulation training (Figure 4). Trainees are required to meet the benchmark criteria in the basic skill exercises before progressing to more advanced skill tasks. Whilst participants will achieve this at different rates, it will ensure that all participants completing the course will have achieved competency in the fundamental skills of robotic surgery.

A number of limitations to this study should be highlighted. It must be remembered that competency demonstrated through achievement of the benchmark score will be specific to the skill and context. Predictive validity for VR robotic simulation has been established but achieving metric scores equivalent to an expert will not imply the trainee as the same clinical performance capacity. Secondly the use of assessment metrics risks participants focussing only on improving their 'score' rather than developing the correct technique. Although using multiple metrics will focus participants on skill areas in which they are deficient, the potential remains for participants to learn only to complete the specific task rather than acquire the necessary psychomotor skills. Mentorship and teaching throughout the course are necessary to avoid such training errors.

Conclusion

Analysis of the HOT course data has provided viable benchmark scores for use during VR simulation training. A benchmark of 75% below the mean expert score offers a challenging but obtainable score for participants to achieve during the HOT course. Based on our analysis we suggest a modular VR training incorporating basic and advanced skill exercises. Continued analysis of HOT results will allow adaptation of these threshold values.

Clear goals set through benchmarking offer objective targets for students and shift training from case volume based training to a more efficient competency based curriculum.

1. Aydin A, Ahmed K, Shafi AMA, Khan MS, Dasgupta P. The role of simulation in urological training - A quantitative study of practice and opinions. *Surgeon*. July 2015. doi:10.1016/j.surge.2015.06.003.

2. Moglia A, Ferrari V, Morelli L, Ferrari M, Mosca F, Cuschieri A. A Systematic Review of Virtual Reality Simulators for Robot-assisted Surgery. *Eur Urol*. January 2014:–. doi:10.1016/j.eururo.2015.09.021.

3. Kumar A, Smith R, Patel VR. Current status of robotic simulators in acquisition of robotic surgical skills. *Curr Opin Urol*. 2015;25(2):168–174. doi:10.1097/MOU.0000000000000137.

4. Ahmed K, Jawad M, Abboudi M, et al. Effectiveness of procedural simulation in urology: a systematic review. *J Urol*. 2011;186(1):26–34. doi:10.1016/j.juro.2011.02.2684.

5. Aydin A, Shafi AMA, Khan MS, Dasgupta P, Ahmed K. Current Status of Simulation and Training Models in Urological Surgery: A Systematic Review. *J Urol*. March 2016. doi:10.1016/j.juro.2016.01.131.

6. Khan MS, Ahmed K, Gavazzi A, et al. Development and implementation of centralized simulation training: evaluation of feasibility, acceptability and construct validity. *BJU Int*. 2013;111(3):518–523. doi:10.1111/j.1464-410X.2012.11204.x.

7. Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. *Med Educ*. 2009;43(1):21–27. doi:10.1111/j.1365-2923.2008.03245.x.

8. Raison N, Ahmed K, Fossati N, et al. MP10-12 Assessment and Validation of the EAU Hands-On-Training Course in Robotic Surgery. In: Vol 29. 2015:P1–A457. doi:10.1089/end.2015.29003.abstracts.

9. Brunner WC, Korndorffer JR, Sierra R, et al. Determining standards for laparoscopic proficiency using virtual reality. *Am Surg*. 2005;71(1):29–35.

10. Abboudi H, Khan MS, Guru KA, et al. Learning curves for urological procedures: a systematic review. *BJU Int*. 2014;114(4):617–629. doi:10.1111/bju.12315.

11. Bland JM, Altman DG. The use of transformation when comparing two means. *BMJ*. 1996;312(7039):1153.

12. Fossati N, Buffi NM, Canda AE, et al. *488 Factors Influencing Performance During Robotic Surgery Training: Results From the EAU Robotic Urology Section HOT-Program*. Vol 14. European Urology Supplements; 2015. doi:10.1016/1569-9056(15)60481-X.

13. Fisher RA, Dasgupta P, Mottrie A, et al. An over-view of robot assisted surgery curricula and the status of their validation. *Int J Surg*. 2015;13C:115–123. doi:10.1016/j.ijsu.2014.11.033.

14. Burke LA, Hutchins HM. Training Transfer: An Integrative Literature Review. *Human Resource Development Review*. 2007;6(3):263–296. doi:10.1177/1534484307303035.

15. Yudkowsky R, Park YS, Lineberry M, Knox A, Ritter EM. Setting mastery learning standards. *Acad Med*. 2015;90(11):1495–1500. doi:10.1097/ACM.0000000000000887.

16. Kane MT, Crooks TJ, Cohen AS. Designing and Evaluating Standard-Setting Procedures for Licensure and Certification Tests. *Adv Health Sci Educ Theory Pract*. 1999;4(3):195–207. doi:10.1023/A:1009849528247.

17. Leung RM, Leung J, Vescan A, Dubrowski A, Witterick I. Construct validation of a low-fidelity endoscopic sinus surgery simulator. *Am J Rhinol*. 2008;22(6):642–648. doi:10.2500/ajr.2008.22.3245.

18. Aggarwal R, Grantcharov TP, Eriksen JR, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Ann Surg*. 2006;244(2):310–314. doi:10.1097/01.sla.0000218094.92650.44.

19. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med*. 2008;15(11):988–994. doi:10.1111/j.1553-2712.2008.00227.x.

20. Fitts PM1, Posner MIJA. Human performance. 1967.

Legends to Figures

Figure 1: Comparison of Participant Performance with Benchmark Score

Figure 2: Comparison of Novice and intermediate Participant Scores to Benchmark Score

Figure 3: Comparison of Novice and intermediate Participant Scores to Benchmark Score

Figure 4: Virtual Reality Curriculum Training Programme for Robotic Surgery

Legend to Table

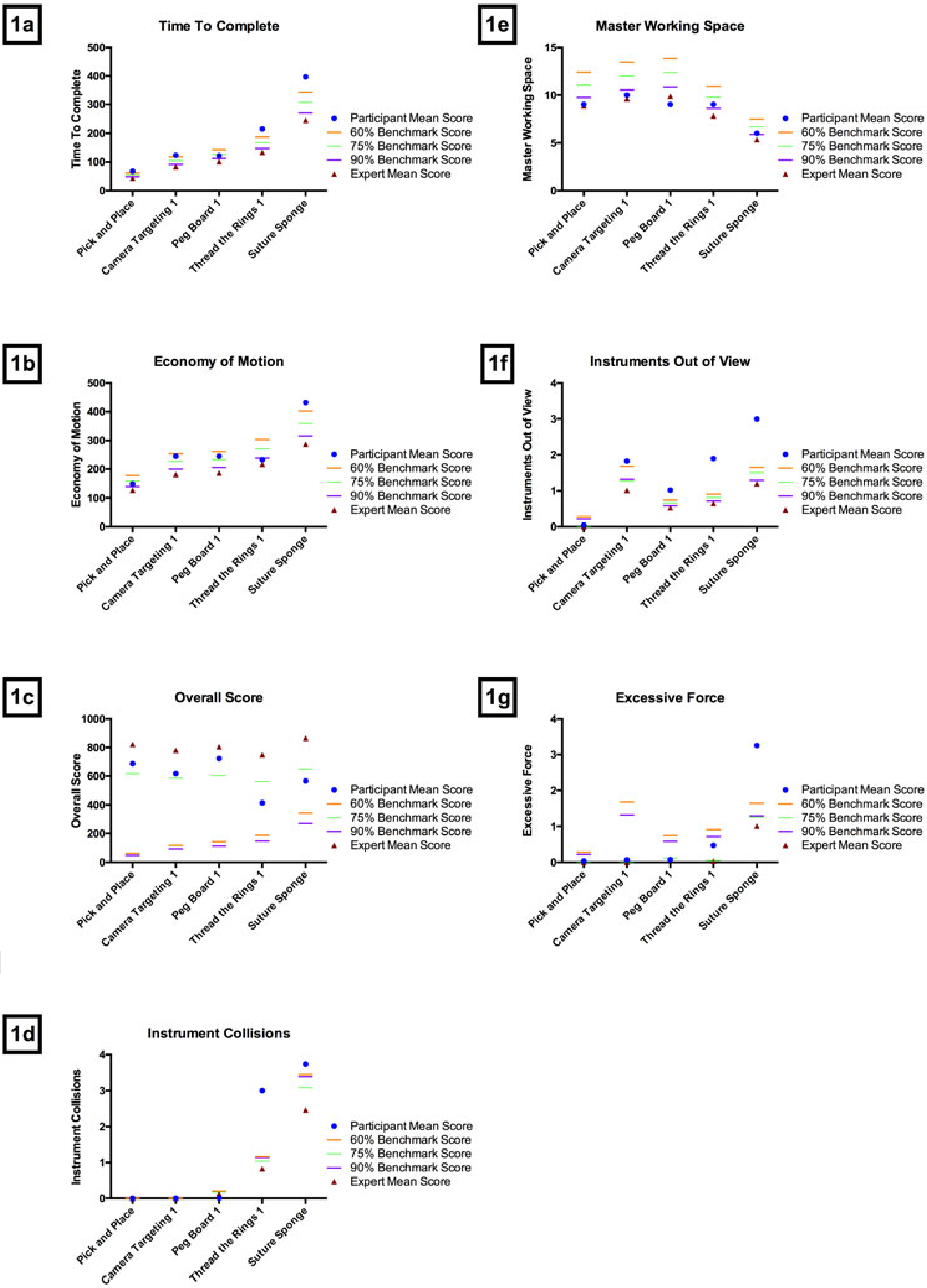
Table 1: Demographic details and clinical experience of HOT course participants

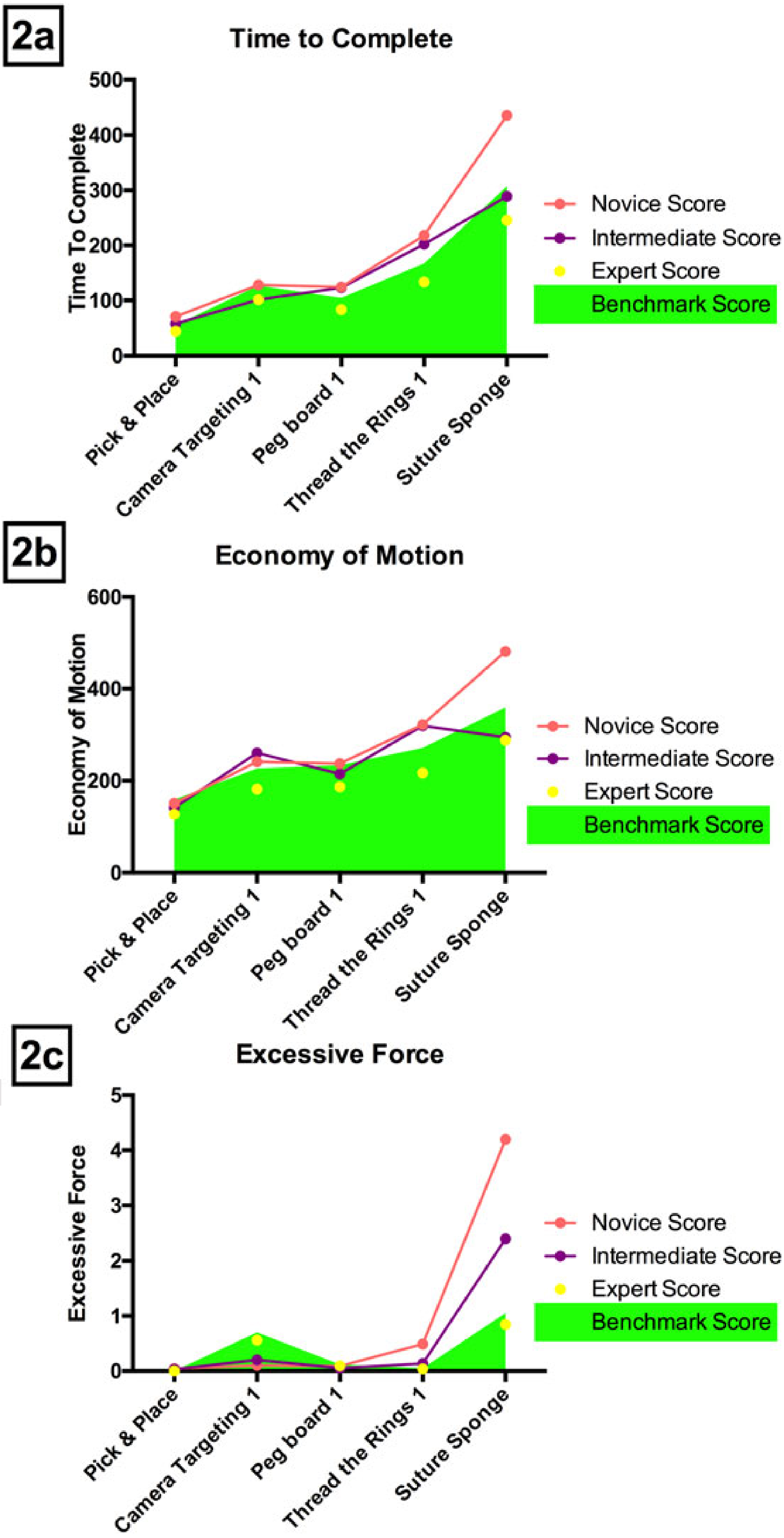
Conflicts of Interest and Financial Disclosures

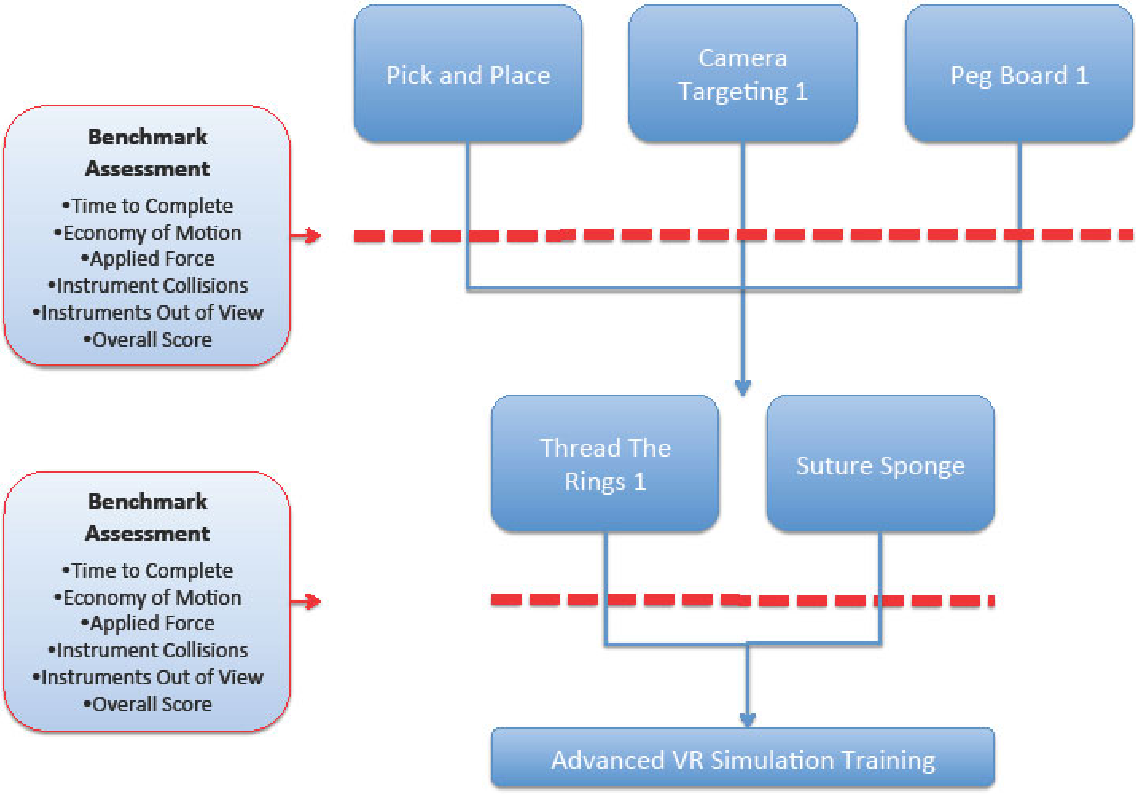
No authors have any conflicts of interest to disclose.

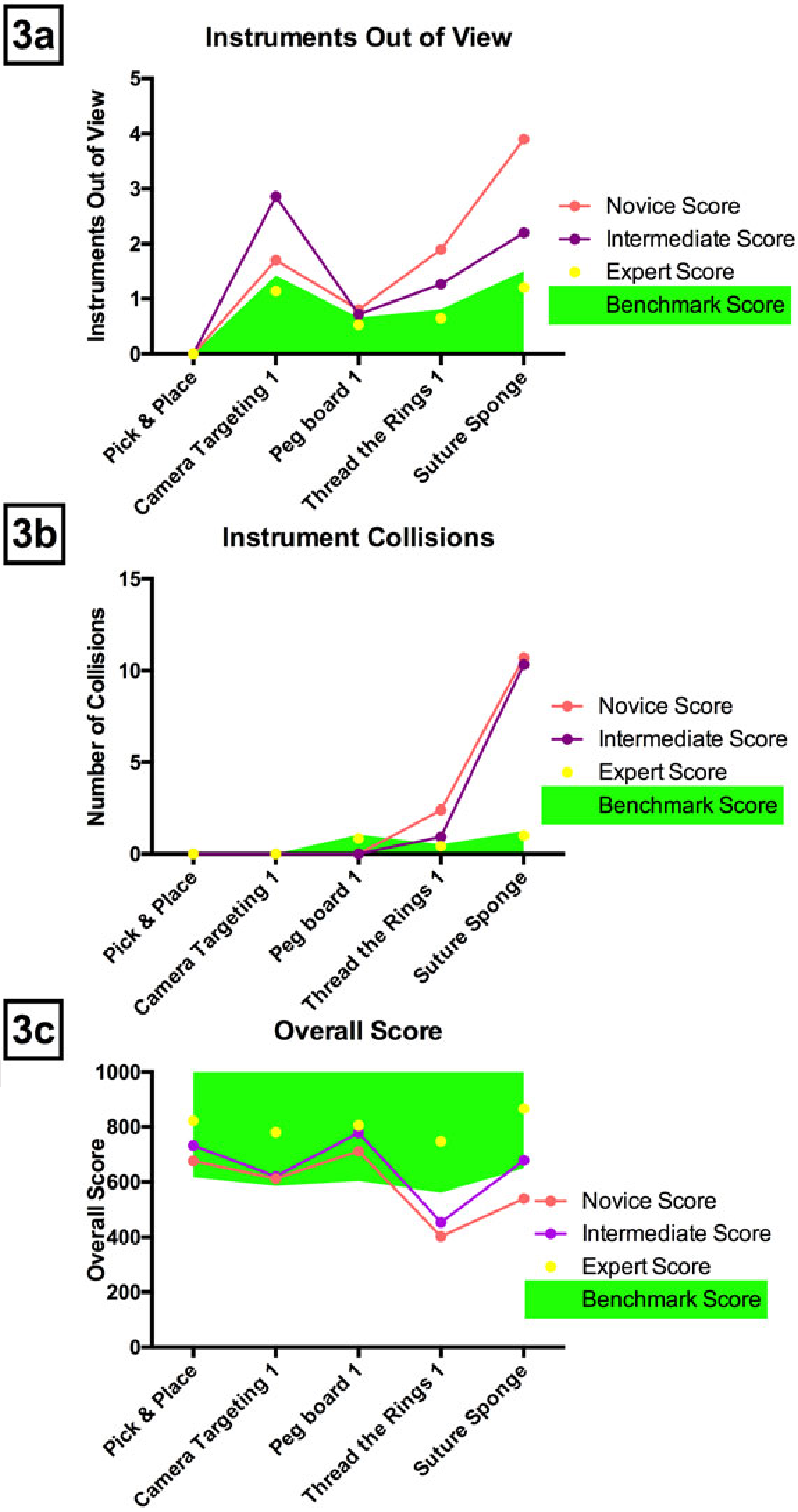
Funding/Support and role of the sponsor: Nicholas Raison acknowledges financial support from the Vattikuti Foundation. Nicholas Raison, Kamran Ahmed and Prokar Dasgupta acknowledge the support of the Medical Research Council Centre for Transplantation and King's College London. Kamran Ahmed and Prokar Dasgupta acknowledge the support of financial support from the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health. Prokar Dasgupta also acknowledges the support of the Medical Research Council Centre for Transplantation, London Deanery, London School of Surgery, Guy's and St. Thomas' Charity, European Union, Vattikuti Foundation, and Olympus.

Age **Mean Years, SD**	38.39 ± 9.48
Gender (%)	75.6% Male : 24.4% Female
Level of Training (%)	45.8% Resident : 54.2% Attending
Robotic Assistance Experience (mean no. of cases, SD)	21.3 ± 52.0
Robotic Surgical Experience (mean no. of cases, SD)	5.56 ± 28.3
Laparoscopic Surgical Experience (mean no. of cases, SD)	68.5 ± 135.1
Robotic Simulation Experience (%)	52.7%


1a
Time To Complete
Time To Complete
1e
Master Working Space
Master Working Space
1b
Economy of Motion
Economy of Motion
1f
Instruments Out of View
Instruments Out of View
1c
Overall Score
Overall Score
1g
Excessive Force
Excessive Force
1d
Instrument Collisions
Instrument Collisions
Pick and Place
Camera Targeting 1
Peg Board 1
Thread the Rings 1
Suture Sponge
Participant Mean Score
60% Benchmark Score
75% Benchmark Score
90% Benchmark Score
Expert Mean Score


2a
Time to Complete
Time To Complete
500
400
300
200
100
0
Pick & Place
Camera Targeting 1
Peg board 1
Thread the Rings 1
Suture Sponge
Novice Score
Intermediate Score
Expert Score
Benchmark Score
2b
Economy of Motion
Economy of Motion
600
400
200
0
Pick & Place
Camera Targeting 1
Peg board 1
Thread the Rings 1
Suture Sponge
Novice Score
Intermediate Score
Expert Score
Benchmark Score
2c
Excessive Force
Excessive Force
5
4
3
2
1
0
Pick & Place
Camera Targeting 1
Peg board 1
Thread the Rings 1
Suture Sponge
Novice Score
Intermediate Score
Expert Score
Benchmark Score


Benchmark Assessment
•Time to Complete
•Economy of Motion
•Applied Force
•Instrument Collisions
•Instruments Out of View
•Overall Score
Pick and Place
Camera Targeting 1
Peg Board 1
Benchmark Assessment
•Time to Complete
•Economy of Motion
•Applied Force
•Instrument Collisions
•Instruments Out of View
•Overall Score
Thread The Rings 1
Suture Sponge
Advanced VR Simulation Training


3a
Instruments Out of View
Instruments Out of View
0
1
2
3
4
5
Pick & Place
Camera Targeting 1
Peg board 1
Thread the Rings 1
Suture Sponge
Novice Score
Intermediate Score
Expert Score
Benchmark Score
3b
Instrument Collisions
Number of Collisions
0
5
10
15
Pick & Place
Camera Targeting 1
Peg board 1
Thread the Rings 1
Suture Sponge
Novice Score
Intermediate Score
Expert Score
Benchmark Score
3c
Overall Score
Overall Score
0
200
400
600
800
1000
Pick & Place
Camera Targeting 1
Peg board 1
Thread the Rings 1
Suture Sponge
Novice Score
Intermediate Score
Expert Score
Benchmark Score